

**Epidemiology and Genetics Research Program
National Cancer Institute
National Institutes of Health
Department of Health and Human Services**

**First Workshop on Cancer Epidemiology and Genomics Variation in
Hispanic/Latino Populations Within the American Continents**

**Marriott Bethesda North Hotel and Conference Center
North Bethesda, MD**

May 4–5, 2006

Patterns of Genetic Variation in Indigenous Populations from South America

Andre Ruiz-Linares, M.D., Ph.D., University College, London, United Kingdom

Dr. Ruiz-Linares made a presentation on the evolutionary history of Native American populations. Unresolved questions include the time of initial entry into the Americas, the migratory pattern from Asia, the routes of dispersal and pattern of differentiation in the Americas, and the consistency of genetic and non-genetic information (e.g., archaeological or linguistic). Greenberg's three migration model was described. This model posits that three major linguistic families correspond to three migrations across Beringia, and that the initial migration led to the development of the Clovis culture (~13,000 years ago) and the Amerind linguistic family.

Genetic data currently available in Native American populations include classical markers (blood groups and proteins), haploid systems (mtDNA and Y chromosome) but limited information for autosomal DNA markers. Five Native American populations were recently included in a worldwide survey employing genome-wide microsatellite screens. Dr. Ruiz Linares described unpublished genetic marker data on 25 native populations and 13 admixed Latino groups (from Mexico/Central and South America). The markers examined were 751 microsatellites and 600 insertions/deletions (InDels).

Analyses of population diversity and diversification in the Americas, particularly Central and South America, was presented. Within and between population diversity analyses included gene diversity, with population-specific F_{ST} , as well as structure analysis, multidimensional scaling and population trees based on Nei's 1983 genetic distance. In addition a tree relating 431 Native American individuals based on the proportion of alleles shared between them was displayed.

Dr. Ruiz-Linares concluded that extensive population structure exists among Native Americans, that there is a North-South gradient in diversity and population structure in the Americas and a contrasting East-West diversity in South America. This, pattern of diversity is likely to reflect colonization routes, gene flow, or patterns of demographic expansion in the Americas. The population tree obtained shows a good correspondence with the linguistic classification of the populations examined. Finally, the differentiation of some of the major linguistic subfamilies of Amerind appears to have occurred in rapid succession.

Population Stratification Confounds Genetic Association Studies Among Latinos

Hua Tang, Ph.D., Fred Hutchinson Cancer Center, Seattle, WA

Dr. Tang noted the concern that population-based case-control association studies in Latino populations may produce false-positive associations as a result of confounding caused by population stratification. For example, cases on average may share a greater degree of European ancestry than controls. As a result, a genetic variant that occurs more commonly in the European population also may occur at a higher frequency in cases. This may mislead researchers to conclude that a variant is causally related to a disease when it is not. Variants that are related to a disease through population stratification may be related only spuriously and may not provide any meaningful information on disease etiology.

Three necessary conditions for confounding are: (1) ancestral allele frequencies differ at the candidate locus; (2) ancestry proportions vary among individuals; and (3) phenotype (disease risk) varies as a function of ancestry proportion. In designing studies, Dr. Tang noted that if any one of the three necessary conditions is excluded, the risk of confounding is low. In addition, if the ancestry variation is known, specific components of ancestry can be modeled or controlled for explicitly.

In conclusion, Dr. Tang suggested that future admixture studies should focus on specific ancestry components, and these components likely will vary based on disease risk/phenotype. There should be a systematic examination of variation in ancestry proportions, genetic divergence among ancestral populations, and the ways in which phenotypes vary with ancestry in admixed populations.

Cancer Incidence and Mortality Trends in Hispanic/Latino Populations in the Americas

Barry Miller, Ph.D., Surveillance Research Program, National Cancer Institute, Bethesda, MD

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (www.seer.cancer.gov) is responsible for the collection and reporting of cancer incidence and survival data from fifteen population-based central cancer registries that cover 26 percent of the U.S. population. The U.S. racial/ethnic population coverage in SEER includes 23 percent of African Americans, 23 percent of Whites, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of native Hawaiian and other Pacific Islanders. SEER covers geographically and demographically diverse populations in the U.S. including all residents of the states of CA, CT, HI, IA, KY, LA, NJ, NM, and UT; metropolitan areas of Atlanta, Detroit, Seattle; selected rural Georgia counties; and American Indian/Alaska Native populations in AK and AZ. The publicly available SEER database contains information on in situ and invasive cancer cases diagnosed among residents of the coverage areas, including a description of the neoplasm (type of cancer, histology, behavior, grade, and extent of the disease), the first course of cancer treatment, vital status, survival time, underlying cause of death, and demographic information (such as age at diagnosis, race/ethnicity, sex, county of residence, and county-level sociodemographic data).

SEER data can be used to derive age-adjusted incidence rates for Latinos/Latinas by gender for diagnoses in 1992 and forward. An analysis of diagnoses in 2000-2003 shows that Latinos have lower overall cancer incidence rates than non-Latino Whites, but a higher incidence of cancers of the stomach and liver/intrahepatic bile duct. Latinas also have lower overall cancer incidence rates than non-Latina Whites, but a higher incidence of cervical and stomach cancers. On a positive note, the cervical cancer incidence rates have been dropping steadily among Latinas and are getting closer to the rates in non-Latina Whites. When age-adjusted incidence rates are partitioned by stage at diagnosis, Latinos have slightly lower proportions of local/regional prostate cancers and localized colorectal cancers and slightly higher proportions of distant stage for these cancers when compared to non-Latino Whites. Latinas have lower proportions of localized breast and cervix cancers and higher proportions of regional stage disease when compared to non-Latina Whites. The percentage of women who received breast-conserving surgery among those diagnosed with early stage disease and tumors less

than or equal to 2 cm are comparable among Latinas and non-Latina Whites. Similar percentages of Latino and non-Latino White men received radical prostatectomy among those diagnosed under the age of 70 with localized or regional prostate cancer.

The SEER Web Site contains further resources, such as cancer statistics, databases, publications, and scientific systems such as SEER*Stat, DEVCAN, and others.

Cancer Epidemiological Studies in Latin Americans

Veronica Wendy Setiawan, Ph.D., University of Southern California, Los Angeles, CA

The Multiethnic Cohort Study (MEC), conducted in Hawaii and Los Angeles, had as its objective the elucidation of the relation of diet, lifestyle factors, and genetic susceptibility to cancer risk. Potential participants were identified through driver's license files from the Departments of Motor Vehicles, voter registration lists, and Health Care Financing Administration data files. The cohort consists of more than 215,000 men and women (aged 45-75 years at baseline) and comprises mainly five self-reported racial/ethnic populations: African Americans, Japanese Americans, Latinos, Native Hawaiians, and Whites living in Hawaii and California (mainly Los Angeles County). The study involved a self-administered mail questionnaire addressing diet, demographic factors, anthropometric measures, personal behaviors (smoking, sun exposure, physical activity), history of prior medical conditions, use of medications, family history of cancer and for women, reproductive history and exogenous hormone use. The incident cancer cases are identified by linkages to the NCI SEER registries, and the cohort was distributed by sex, ethnicity, and migration status. Analyses were performed for Y-chromosomal haplogroup distribution in the MEC, as well as smoking and drinking, obesity, and vigorous physical activity by sex and ethnicity.

The top five cancers among Hispanics found by the MEC were prostate, colorectal, lung, stomach, and liver for men, and breast, colorectal, uterine corpus, lung, and ovary for women. For prostate cancer, the MEC examined nearly 5,000 cases for age-adjusted prostate cancer incidence; Latinos were second among the population groups. Risk factors were delineated as age, ethnicity (with the highest risk among African Americans), and family history of prostate cancer. Associations with dietary and other lifestyle factors remain unclear.

More than 3,000 cases were analyzed regarding breast cancer incidence. Risk factors included ethnicity, with the greatest risk for Native Hawaiians and Japanese Americans. There were consistent associations with established risk factors as well, including early menarche, late age at menopause or first birth, nulliparity, high body weight, use of alcohol, and hormone replacement therapy use.

The MEC conducted an analysis based on these risk factors by racial/ethnic group and examined predicted versus observed relative risks of breast cancer. It was reported that breast cancer risk factors fully explain the lower breast cancer rates among U.S.-born Latinos but only partially among foreign-born Latinas. The finding was consistent with other studies. This suggests the importance of other unknown factors, such as exposures early in life, habits maintained among immigrants, or diet. The effect of dietary fiber intake on estrogen levels also was analyzed; higher levels of dietary fiber intake were associated with lower estrogen levels in postmenopausal Latino women.

The age-adjusted incidence of colorectal cancer was also studied by the MEC, with African American women and Japanese men appearing to have the highest risk for the disease. High dietary fiber intake was associated with decreased risk and smoking with increased risk.

The MEC's study of age-adjusted incidence of lung cancer revealed that the highest risk was for African American men and women. Several other analyses, including smoking prevalence by ethnic group and gender, the quantity of cigarettes smoked each day, and ethnic differences in association between smoking and lung cancer, found substantial ethnic differences in lung cancer risk associated with smoking. Specifically, African Americans and Native Hawaiians have relative risks three to four times higher than those of Japanese and Latinos and twice than those of Whites at lower levels of smoking. Moreover, African American and Native Hawaiian smokers may be more susceptible to the carcinogenic effects of cigarette smoking at lower smoking doses.

The MEC studied almost 325 cases of endometrial cancer in postmenopausal women and determined that Whites were at greater risk. Risk factors were defined as early menarche, late menopause, nulliparity, estrogen therapy use, and obesity. Long-term oral contraceptive use is protective against endometrial cancer, whereas obesity is the strongest risk factor.

Genetic association studies in the MEC include: candidate gene studies in breast, prostate, colorectal and endometrial cancers, which involve resequencing and haplotype-based analysis of sex-steroid, growth factor, and DNA repair genes; the measuring of hormones (sex steroids, insulin-like growth factor, and prolactin; admixture mapping in African Americans (prostate cancer); and planned whole-genome association studies in breast and prostate cancer.

Genetic Epidemiology in Latino Populations

Neil Risch, Ph.D., University of California at San Francisco, San Francisco, CA

Dr. Risch provided background on genetic variability among Latinos and challenges associated with studying Latino populations. These include the heterogeneity of the population, which is culturally and genetically very diverse, and the fact that Latinos often are part of a complex admixed genetic ancestry. One consequence of this complexity is that it may be difficult to appropriately match cases and controls in genetic association studies. This improper matching could lead to biased association study results. In addition, researchers have not been able to settle on a definition of the terms "Latino" or "Hispanic." As the Latino population continues to grow in the United States, these questions need to be addressed to address the health needs of this population.

To illustrate the challenges in conducting research in Latino/Hispanic populations, two epidemiologic studies were summarized. The Family Blood Pressure Program (FBPP) examined genetic and environmental determinants of hypertension in families and found that study results are dependant on the quality and quantity of the markers, and that more thorough study occurs when working with a larger number of markers. Genetic cluster analysis indicated concordance with the self-identified race/ethnicity of participants. However, admixture analysis also pointed out the mixture of continental ancestries within individuals among Latinos and African Americans. Similarly, there was strong confounding between genetic and nongenetic factors when examining group differences in prevalence of diseases or traits. The FBPP results illustrate that genetic,

social, cultural, and economic factors underlying group distinctions are highly confounded and one of the major problems in performing group comparisons.

The second study, the case-control Stanford-Kaiser Cardiovascular Disease Project (SCDP), included nine self-identified race and ethnicity groups. Participants in the SCDP also could choose more than one race or ethnicity group and were asked about their grandparents' ancestries and countries of origin. The SCDP also was a candidate gene study, with approximately 77 candidate genes; approximately 467 single nucleotide polymorphisms (SNPs) of the candidate genes have been genotyped. Analysis in the SCDP included the study of a single gene with linked disequilibrium and was based on haplotype frequencies; results indicated that those self-identified as Hispanic or South Asian were difficult to distinguish genetically. The explanation for this result was not clear, but this information may be relevant to cancer epidemiology if the influence of genetics versus environmental etiology for specific cancers can be related to genetic or haplotype differences.

Low-frequency genetic variants tend to be specific to populations, continents, or sometimes an ethnicity within a continent. The degree to which genetic variation between racial and ethnic groups contributes to the differences and prevalence of common or complex traits remains incompletely understood. Analytic tools, such as multigenic models, have been developed to determine the extent of genetic contributions to group differences. In a polygenic threshold model, analyses can address how differences in heritability in various populations might affect the degree of differences in the prevalence of a gene; of particular interest is the consideration of relative risk for disease associated with one or more susceptibility genes and how this would differ between population groups.

Dr. Risch described a specific analysis using a model that was conducted to show the heritability of genetic factors and how they varied little between population groups (usually less than 3 percent). The SNP allele frequency differences had little impact on differences in heritability. In terms of the relative risk associated with the contributing gene, differences in heritability depended primarily on how much variance was contributed by that gene. Even for common traits and common variation, including genetic variation, the amount of differentiation apparent between groups, especially for high heritability traits, can generate some group differences comparable to what might be seen in the population as well as disease prevalence differences.

A study of mortality data from the U.S. National Center for Health Statistics and the Centers for Disease Control and Prevention indicated that recent Hispanic migrants to the United States appear to have lower rates of some major diseases, such as heart disease, cancer, and respiratory disease, than Native Americans and are at lower relative risk. Hispanics do show intermediate risk for *diabetes mellitus* and stroke, where their risk is much closer to Whites. Although genetic factors could be determining differences, Dr. Risch said that there is not an obvious genetic pattern that can explain these differences, at least from an ecological view. Although Hispanics are genetically admixed between multiple ancestral populations, these disease rates are not intermediate between (for example) Europeans and Native Americans. Thus, differing rates of cancer in ethnic populations are not strongly suggestive that genetics completely explains the inter-individual variability in disease risk. For example, Latinos have higher cancer rates at some major cancer sites than Asians in the US, and but many rates are lower in Latinos than in Whites. If genetic variations were underlying these group

differences, a random picture would develop. These data suggest major environmental factors may influence cancer rates more than genetics.

Admixture analysis also can be employed to distinguish between genetic and nongenetic sources of population differences. The power of this analysis depends on the variation in admixture levels that exist within and between populations. One factor that contributes to variation is the recency of the admixture. In the admixture analysis for the FBPP, the blood pressure level correlation between African American and Mexican American individuals was examined. Differences were fairly modest between hypertensives and normotensives, although hypertensives tended to have more African ancestry than normotensives. There was, however, a significant association of body mass index with African ancestry, and overall, the study indicated heterogeneity of these relationships within these populations.

Admixture mapping has been available for approximately the last decade, and studies about this process have begun to appear. The power of this method for gene mapping depends on the magnitude of difference in allele frequencies between the ancestral populations that contribute to the admixture group. If these allele frequency differences are not substantial, the ability of admixture mapping to detect genetic effects will be low. Based on simulations discussed previously, and knowing that the allele frequency differences between ancestral groups on average is not large, it is not known how many genes will be identified through admixture mapping.

A recent study published in the *New England Journal of Medicine* on lung cancer rates and group differences demonstrated an interaction between ethnicity and smoking on lung cancer rates. Rates are increased in African Americans, especially males; lung cancer rates also are increased in Pacific Islander males but are decreased considerably in Latinos and Japanese Americans. In people who never smoked, the relative risks are highly attenuated, suggesting that there is an interaction between ethnicity, smoking, and lung cancer rates. It is unclear whether this interaction is based on genetics, the environment, or some other factors.

Preliminary results of recent analyses on asthma and Puerto Ricans indicate that the relationship between ancestry variation and environmental factors in disease etiology is complex. Asthma currently is an epidemic disease in Puerto Rican populations, much more significantly than any other minority group in the United States. The minority group within the United States with the lowest rates of asthma is Mexican Americans, even though these two groups—Mexican Americans and Puerto Ricans—probably are the most genetically similar. If genetics are playing a role in these differences, the relationship is not linear according to Dr. Risch. He noted that ancestry here can be a surrogate for some other factor.

Potential of Network Collaborations

Amelie G. Ramirez, Dr.P.H., Baylor College of Medicine, Houston, TX

Redes En Acción (Networks in Action) is composed of behavioral and community-based researchers across the United States and is funded by NCI's Centers to Reduce Cancer Health Disparities (CRCHD). The project grew out of Requests for Applications (RFAs) issued to establish leadership initiatives on cancer, particularly to focus on cancer prevention and control in special populations. These RFAs also emphasized training

and community education and gave rise to community network programs such as *Redes En Acción*.

Redes En Acción mentoring efforts encourage Latinos to enter and remain in research fields, beginning at the undergraduate level. More than 130 individuals have been mentored by the network and gained experience in grant writing, field research, data analysis and reporting, and manuscript development. *Redes En Acción* also has mentored several junior faculty members applying for pilot funding from NCI. Over a 5-year period, *Redes* achieved one of the highest success rates (55%) for community-based, pilot projects submitted by junior investigators who had not been funded in the past, the majority of which have gone on to receive independent funding of more than \$9 million. The network also participates in extramural research efforts to expand research in the regions. *Redes En Acción* co-PIs are involved in more than 80 different research projects that leverage \$27 million for the network and have resulted in more than 200 scientific articles.

Redes En Acción initiated a national program, including a national media campaign, to increase awareness of clinical trials in the Latino community and increase participation of underrepresented minorities in NCI's clinical trials. One finding of this effort was a recognition that the Latino community reacts negatively to the word "trial," with "studies" being a term that had a more positive connotation. *Redes En Acción* also produces quarterly newsletters that focus on network efforts and highlight Latino researchers as role models for the community. Network personnel have participated in 1,400 community and professional events and have developed a popular Web site.

Dr. Ramirez cited the Breast Cancer Genetics Survey Project as an example of a *Redes En Acción* network collaboration, funded in part by NCI. Because little information existed about Latinos and other minority populations and their attitudes and opinions about genetic testing for breast cancer, *Redes* approached the Susan G. Komen Foundation to help bring together key population groups and design a culturally sensitive survey to be used to assess similarities and differences in breast cancer genetics knowledge, attitudes, and behaviors among African American, Appalachian, Asian American/Pacific Islander, and Native American/American Indian women and Latinas. This was the first time that different population groups were brought together to collaborate on a genetics cancer research project. The project developed a core questionnaire with items common to all five populations and culturally specific supplemental questions developed for each group. Areas addressed included breast cancer and genetic testing history, knowledge, attitudes and behaviors, and demographics. The collaborating groups are independently seeking funds to administer the questionnaire within their populations. *Redes En Acción* has received funding from the Komen Foundation to implement the Latino survey with women in South Texas during the summer of 2006.

The Hispanic/Latino Genetics Community Consultation Network Project is another network collaborative initiative that was established to enhance the knowledge of basic genetics in the Latino community. The goal was to convene a national summit of key Latino investigators to develop recommendations for research, healthcare services, professional education and training, and public education and outreach on genetic issues confronting the U.S. Latino community. More than 200 representatives from throughout the United States were involved. A consensus report is available at www.redesenaccion.org, and results have been published in *Cancer Research*.

Recommendations included using the report to guide research and policy decisions on genetics and conducting consensus dialogue with other ethnic groups.

A third example of a network collaborative project cited by Dr. Ramirez focused on testing three methods for recruiting Latinos into a cancer genetics registry. NCI had noted that the registry contained few minorities and few Latinos and asked *Redes* to help recruit more minorities into the database. The three methods tested by *Redes* were direct mail; direct mail plus bilingual materials; and direct mail, bilingual materials, and telephone contact, which yielded the best results. Recommendations that resulted from this project include comparing recruitment strategies, considering cost-effectiveness analyses, and increasing research on recruitment strategies for different ethnic groups.

The *Redes En Acción* experience has shown that well-managed networks can provide significant benefits, but efficient communications must be maintained, goals and objectives must be co-created, and leadership must be strong and centralized. Good networks can help foster new alliances, stretch resources, and accelerate change. The genetics research community should realize how critical such collaborations are—not only across programs, but across disciplines and continents, as well.

Ethics, Identity and Social Justice

Vivian Ota Wang, Ph.D., Senior Advisor – Office of Behavioral and Social Sciences Research (OBSSR), Office of the Director, and Program Director - Ethical, Legal, and Social Implications Research Program, National Human Genome Research Institute, Bethesda, MD

When discussing ethics in genetics, one is dealing in a social context in which fairness and equity should play large roles. There are ethical, legal, social, and behavioral implications in this context which include: issues surrounding nature versus nurture, privacy and confidentiality, protections for research participants and informed consent, intellectual property, and human variation, among others. Dr. Ota Wang posed three questions: (1) What do you care about versus what do you not care about? (2) What do you see versus what do you ignore? (3) Who are you versus what do others think you should be? She observed that much of the work occurring in the admixture field addresses the third question and that attendees of this conference are making assumptions about the study participants. A topic of prime concern is whether race really matters. Although the word “race” is contentious, Dr. Ota Wang posited that it is important to talk about what race means in the context of population differences. She noted that the workshop presentations have discussed more of how to understand populations and less of the implications of the former.

Dr. Ota Wang presented different models in how race and racial categories are understood including Linnaeus’ Classification System, Critical Race Theory, Racial Identity Theory and clines and population migration history. She showed a chart illustrating within group differences are greater than between group differences with four overlapping Gaussian curves representing Asian, White, Hispanic, and Black population groups. She then discussed Linnaeus’ *Systemae Naturae* (1758) human racial classification system (Europeus, Americanus, Asiaticus, and Africanus) as his attempt to understand his experience of population differences by descriptively categorizing groups of people by skin color (white, reddish, sallow/yellow, and black, respectively), morality, and personality traits. She remarked that the zeitgeist of the period in history and moral issues influence value judgments. She shared early intelligence tests,

conducted with the “state-of-the-art” science that existed in the early 1900’s that used the slope of an individual’s facial profile as a measure of intelligence with people with lower intelligence having a greater degree of slope (e.g., people from the African continent).

The critical race theory offers another way in which people consider race and population. It posits five ideas: (1) race is not equal to ancestry, skin color, eye shape, hair texture, or other physical characteristics; (2) race is not genetic; (3) race is beliefs about ancestry, nationality, language, religion, skin color, and other phenotype information; (4) racialized beliefs are relational and created by interactions based on and reinforced by a person’s beliefs about ancestry, nationality, and so forth; (5) race is historically and geographically fluid.

Racial identity theory applies to all racial groups and involves dual perspectives: self and own racial group and self in relation to other groups. There are implications regarding information processing; perceptions, attitudes, and behaviors; and representative sampling and research design. Dr. Ota Wang recalled the universal principles of genetics in that all people are human first and that nearly all diseases (except some cases of trauma) have a genetic component; in addition, everyone carries a significant number of DNA glitches.

Although races are not distinctive, natural categories, group differences may be used in epidemiology to make statistical predictions. In addition to genomic and environmental factors discussed in literature, social and behavioral factors must be considered. Behavioral factors might include risk perception and attribution as well as decision making. When this workshop provides recommendations, it should consider the validity and reliability of how investigators use and distinguish populations and or racial group memberships including Ancestral Informative Markers; investigator-inferred, self-reported recent ancestral geography; and biobehavioral and genetic markers.

Dr. Ota Wang summarized the choices that researchers can make regarding race. One choice is to abandon race as a variable in biomedical and health equity research and clinical practice. Another choice is to use race in biomedical and health equity research and clinical practice; there are useful ways, including genetics and epidemiology, and genetic data, that one can adjudicate the reality of race and validate racial (and ethnic) categorizations for research and public policy. She suggested that, regardless of what level one was working at—individual, communal, or large-scale population—one needs to keep the larger picture in mind. When conducting population or genetic work, she cautioned against conferring biological or genetic determinants, to use clear and consistent criteria in how group categories are used and described, and use more proximal variables in research designs rather than defaulting to simplistic racial or population labels.

Admixture Mapping for Breast Cancer in Latinas

Elad Ziv, M.D., University of California at San Francisco, San Francisco, CA

Admixed populations are those in which groups with genetic differences have mixed, such as Latinos and African Americans. Studies of admixture generally measure individuals’ ancestry and associate individual ancestry with phenotype. Data from SEER can be used to delineate rates of breast cancer among different populations in the United States. Dr. Ziv and Dr. Esther John estimated genetic ancestry in a case-control

study of breast cancer and ancestry among Latinas using 44 ancestry informative markers. Information on nongenetic risk factors was collected by questionnaire. The investigators found significant differences in genetic ancestry among women from different regions of Latin America, but also considerable variation among women from the same region. There were also associations between ancestry and several reproductive and hormonal factors. Ancestry was not different overall among cases and controls; however, among younger, premenopausal women, risk appears to increase as percent Native American ancestry.

Limitations in the studies include possible bias in recruitment, unmeasured environmental factors, and the limited number of markers used to estimate ancestry.

Genomic Regions Exhibiting Positive Selection Identified From Dense Genotype Data

Joshua Akey, Ph.D., University of Washington, Seattle, WA

Dr. Akey stated that positive selection involves the differential contribution of genetic variants to future generations. Identifying targets of selection provides insight into mechanisms of evolutionary change, clues to evolutionary history, identifying functionally important regions, and mapping complex disease genes. However, unambiguous inferences of selection are difficult.

Positive selection imparts “signatures” on patterns of genetic variation that may include reduced variation, an excess of low frequency alleles, an excess of high frequency derived alleles, increased LD, and increased population structure. Selection is difficult to detect because evolution is a noisy process and because signatures of selection can result from additional perturbations to the standard neutral model.

The population genomics approach attempts to address these issues. This approach holds that genetic drift equally affects all loci in a genome, whereas natural selection acts only on specific loci. Sampling many unlinked regions can help to disentangle the effects of drift and selection. Two complementary approaches have arisen. One approach is to contrast patterns of genetic variation between different classes of sites (e.g., between synonymous and nonsynonymous sites); the second approach seeks to identify outlier loci by examining genome-wide effects and locus-specific effects and looking for unusual patterns of variation.

Dr. Akey and colleagues have used data from the HapMap project (described in Dr. Wall’s presentation) and the Perlegen project (1.58 million SNPs genotyped in 71 African American, Chinese American, and European American individuals). These datasets provide the necessary resources to systematically interrogate patterns of human genomic variation.

A recent analysis compared the distribution of population structure (measured by the summary statistic F_{ST}) across different types of sites. The distribution of F_{ST} has been shown to vary across different functional classes of SNPs, which has been interpreted as differing selective pressures across these functional classes. Dr. Akey and colleagues were interested specifically in testing the hypothesis that positive selection has promoted an enrichment of high F_{ST} values in particular functional categories of genetic variation. Their findings were consistent with the action of local adaptation at least in part helping to shape patterns of nonsynonymous variation between populations.

It is crucial to be careful when interpreting such patterns because the patterns may be observed for reasons that do not relate to selection.

Dr. Akey and colleagues used Perlegen data to examine the second approach, involving gene-based, genome-wide scans based on outlier approaches. The Perlegen data have been shown to have lesser levels of ascertainment bias. For each gene selected for study, the researchers calculated a summary statistic, termed TDGen. This resulted in the identification of specific “candidate selection genes” to be studied in greater detail. Outlier loci were defined as falling in the bottom 1 percent of the empirical distribution of the samples. As a result, 141 candidate selection genes were identified among the African American sample, 130 in the Chinese American sample, and 135 in the European American sample. Forty-one percent of these genes were found to have been shared between two or more of the populations. The candidate selection genes also often were found in clusters, defined as two or more contiguous candidate selection genes. These clusters can be attributed to genetic hitchhiking. These results were found to be in line with those of other studies performed to validate candidate selection genes.

In summary, Dr. Akey noted that he and his colleagues have identified several hundred genes that are consistent with the hypothesis of positive selection. These results were achieved with fairly simple techniques; the opportunity exists to develop more sophisticated approaches that will take more of the data into account. These techniques must be computationally practical and able to address the issue of ascertainment bias. Genome-wide analyses are a beginning and not an end; such studies must yield to more focused, single-locus studies that eventually incorporate phenotypic data as well as functional data. In addition, positive selection might have important implications for disease-gene mapping.

Global Pharmacogenetics Research Networks

Howard L. McLeod, Pharm.D., Washington University School of Medicine, St. Louis, MO

The Pharmacogenetics Research Network (PGRN) and the Pharmacogenetic for Every Nation Initiative (PGENI) are important resources in the field of global health pharmacogenetics. Current therapies are successful in controlling the disease or symptoms of interest in less than 50 percent of patients, in part because most drugs are developed in White European patients with little thought given to how drugs will be used throughout the world. Using information gleaned from the Human Genome Project, better understanding of the genetic basis for “ethnic differences” should help improve disease diagnosis and selection of therapy and offer a way to better integrate medications into national formularies in a safe and effective manner.

The NIH PGRN is based in the USA and focuses on development of robust knowledge on the application of genetics to optimize drug therapy. PGENI will ultimately be active in 104 countries, which contain 78 percent of the world’s population. The objectives of the network are to: (1) promote the integration of genetic information into the public health decision making process; (2) enhance the understanding of pharmacogenetics in the developing world; (3) provide guidelines for medication prioritization for individual countries using pharmacogenetic information; and (4) help build local infrastructure for future pharmacogenetic research studies.

The Initiative has developed a study plan to identify common ethnic racial groups, collect blood samples from each ethnic group, genotype for variants of interest, and generate recommendations for medication selection. The network focuses on systemic drugs from The World Health Organization's (WHO) Essential Medicines List (www.who.int) and has conducted text mining for metabolism, transport, and drug target proteins and allele frequencies of key SNPs in key genes. For example, drug metabolism is affected by ABCB1 genotype, which differs globally. Optimal selection of HIV drugs differs according to ABCB1 genotype. Information concerning genotype of ABCB1 and other genes involved in drug metabolism can help identify population subgroups at higher risk for toxicity or treatment failure and also can be used to prioritize treatment selection from among WHO-recommended therapies. Ethical considerations for public health pharmacogenetics include consultation with communities, the development of a clear mechanism to integrate information, and implementation of safeguards for "genetic orphan" populations.

At present, PGENI is not involved in population genetics research, conducting clinical trials, or performing gene-outcome studies related to pharmacokinetics, toxicity, and efficacy. Lessons learned from PGENI efforts include the importance of involving local health ministries, engaging local stakeholders (e.g., health and community leaders), and ensuring local involvement in the selecting of drugs, inclusion of ethnic/racial groups, and the ethics procedures.

What's in a Word? Models and Realities Underlying the Term "Admixture"

Joanna Mountain, Ph.D., Stanford University, Stanford, CA

The idea of race as a genetic construct is controversial; most geneticists contend that genetic markers show that this is no "pure" race and any classification of races is therefore arbitrary and imperfect. Analyses of haplotypes show that human races are not distinct lineages. Admixture analysis involves the mapping of genes for traits and diseases that have different risks in two or more populations that have admixed recently to form a third hybrid population. The term "admixture" may be problematic in that it evokes the idea of purity and then mixture.

Although "hybrid vigor" is popularly assumed to be beneficial, there also is a belief that advantages may accrue to those choosing an optimal degree of genetic similarity in their (human) mates; optimal fitness may be achieved by selecting a mate who is similar genetically, yet unrelated. This belief does not represent the mainstream attitude, although it has not been completely marginalized; thus, the lay public may believe that admixture is negative. Because the United States, however, has a unique political and social history, including a history of subdivision between ethnic groups, differences among these groups are emphasized. The public also may fear that the study of genetic differences will create stigmatized populations, lead to genetic discrimination, or may reinforce old prejudices, making it difficult to address the issue of genetic differences between different human populations.

Admixture can be thought of as a composite gene pool in which at least some individuals can trace ancestry to more than one population or as the formation of a new population by interbreeding between individuals from genetically divergent parent populations. A key aspect to these definitions is the existence of genetic divergence between parent populations. Admixture mapping requires a measurable distance between parental populations in the frequency of disease-causing alleles. A set of informative markers

that are distributed across the genome relatively evenly also are needed. Complete population isolation is not needed; allele frequency differences can be developed while exchanges occur between groups.

Genetic clustering of 12 populations comprised of 203 individuals shows a pattern of genetic differentiation that gives rise to three groups from the continent of Africa, one oceanic population, and groups from the Americas, Europe, and Asia. Mitochondrial-inferred migration shows populations moving out of Africa, initially to Oceania, into Europe, and eventually into the Americas. These analyses revealed a geographic element to human genetic diversity; people tend to marry or reproduce with those living nearby, which has led to some structuring within the human species.

When the structure within the human species is measured, average F_{ST} (i.e., the divergence measured among populations) has hovered around 0.15 or less for decades; but it is important to note that, when analyzing multiple loci, there is a real distribution of F-statistic (F_{ST}) estimates of the divergence between groups. F_{ST} , and allele frequency differentials are all ways to consider differences between groups. For example, for an F_{ST} of 0.5, there is an F (i.e., ancestry information content) value of 0.3 and a mean allele frequency differential of 0.46. Most allele frequency differentials are lower than 0.46, but this is typical for ancestry informative markers, where large differentials are desired. There are theoretical expectations of how F_{ST} increases, both as a function of time and aspect of population size.

For admixture mapping, the population divergence that provides allele frequency differences is needed, as well as a space of time since the initial admixture, to eliminate disequilibrium between unlinked loci but not for linked loci. Parental populations also should provide adequate genetic breadth. Continued gene flow in both directions is acceptable, although for admixture mapping unidirectional gene flow is optimal.

To allay fears and concerns about admixture mapping, geneticists and epidemiologists can emphasize: (1) the complexity of human history, which has generated the current patterns via the role of geography and geographic distance; (2) the idea that genetic exchange is compatible with these models of admixture; and (3) the possibility of generating genetic differentiation, despite low levels of isolation between populations and relatively recent isolation.

The African Diaspora

John Thornton, Ph.D., and Linda M. Heywood, Ph.D., Boston University, Boston, MA

Voluntary and involuntary migration from Africa to the Americas, often referred to as the African Diaspora, occurred from the early 17th century into the early 19th century. Much of the available information concerning this event comes from merchant ship logs cross-referenced with port records; in 1999, a uniform database to consolidate the records was produced under the auspices of the DuBois Institute at Harvard University. Although not all voyages were recorded, as many as 27,000 were identified, many of them through multiple sources, and this number was later expanded to 34,000 trips.

The shipping records identify four waves of arrivals into different regions of the Spanish Americas from various regions of Africa, resulting in populations that vary according to origins and times of arrival of the Africans. The first wave occurred between 1540 to 1560, from the Senegambian region in West Africa, and went to the large islands of the

Caribbean, especially Santo Domingo (i.e., the Dominican Republic), and then later to Mexico and Peru. A second wave occurred in the early to mid-17th century, almost entirely from Angola, as a result of wars conducted by the Portuguese against the African population during this time period. In the 1640s to 1650s, slaves were imported primarily from English and later Dutch sources, and went mostly to Mexico, Peru, and Columbia. The third wave supplied slaves to the cocoa industry in Venezuela. The final fourth wave began in the late 18th century, conducted mostly under English auspices to provide slaves for the sugar industry in Cuba.

Each wave of migration represented certain areas of Africa, as evidenced by the ethnic names assigned to the arrivals based on their point of origin in Africa; for example, those from Nigeria might be called Ibo or Ibebios. Alonso de Sandoval, a Jesuit priest, published a description of Africa and its ethnography in 1627 based on interviews with slaves and ship captains. This guide contained geographical information that allows the names in Spanish legal documents to be matched precisely with African locations. In the 18th century, Oldendorf, a Moravian missionary who worked among the slaves in the Danish Virgin Islands from 1766 to 1767, interviewed slaves, collected ethnographic and geographical information, and included language samples for each of the nations. Koelle, a German missionary linguist, provided a similar description for slaves brought to Sierra Leone by the British antislavery squadron. Combining shipping data, statistical base, and coastal divisions with the ethnic information helps to develop a picture of the origin of the populations in the various Spanish colonies, which was greatly influenced by the commercial context and routes of the suppliers.

A study of ethnicities on several estates for the years 1544 to 1550 showed that 80 percent of the people mentioned were from West Central Africa, with 13 percent from older Senegambia and Lower Guinea. The African ethnicities identified in the notarial records (i.e., inventories and wills) use ethnic names from the West Upper Guinea coast, such as Wollof, Bran, Mandinga, and Hula, and from West Central Africa. These ethnicities comprise the founder generation of the Afro-Mexican and Afro-Peruvian populations.

The 1570 census for New Spain indicated an African population of 20,569 and revealed the emergence of a mixed population, including Africans born in the Americas and people of African descent. The ethnic makeup of Peru's African population was similar to New Spain prior to 1600 in that a predominance of ethnicities (approximately 74 percent) were from the Upper Guinea coast. In Peru during the third and fourth migration waves (1639 to 1690), 80 percent of 676 Africans came from Angola/Congo and 15 percent from other West African origins. After 1700, when the British dominated, the ethnic origins of slaves in Spanish America changed dramatically, as did their destinations. Few of the slaves went to Mexico, Colombia, Peru, or Venezuela, instead going to the islands of the Caribbean, particularly Cuba. Moreover, the English slave trade focused more on West Africa, and less on the Upper Guinea coast, as indicated by a contract with the English South Sea Company to deliver 4,800 slaves per year to Spanish territories between 1710 and 1739; the origins of these slaves were the Bight of Benin (1,900); Gold Coast (Ghana) (1,500); Gambia (700); and the area from the Gold Coast to Sierra Leone (500 and 200, respectively).

After 1807 when the British officially abolished the slave trade, there was a rapid increase in the number of African-born slaves going to Cuba. Cuban shipping data from the Dubois database for 1776 to 1800 record the arrival of 38,000 slaves and indicate

the points of coastal origin of approximately one-half of them. Forty percent or more came from the Bight of Biafra (Ibo area of Nigeria); 13.5 percent came from the Gold Coast (Ghana); and 12.9 percent, the Bight of Benin. Thus, 67 percent of the slaves came from the area that today spans from the Ivory Coast to Nigeria, with only 23 percent coming from West Central Africa. Information concerning the origins of slaves and the different migratory patterns that brought them to the Americas, along with data concerning intermarriage among slaves, Europeans, and Indians, must be considered to understand the makeup of admixture groups and in studies using haplotype analysis.

Moving Beyond Continental Admixture: What Can Be Said About Intracontinental Genetic Contributions

Mark Shriver, Ph.D., The Pennsylvania State University, University Park, PA

Dr. Shriver focused his talk on genetic issues within a continent, noting that a subset of informative markers is the desired outcome. He posed several questions, including: (1) How can those markers be identified? (2) What is the plan for confirming that these markers are useful? (3) How can one specifically ignore continental-level admixtures if intercontinental stratifications are to be explored? (4) How can one look specifically at intercontinental or ask international questions when a person also has admixture? (5) How can one adjust specifically to the stratification?

Dr. Shriver described a population genomics model. A genome is comprised of thousands of independent parts, explaining why individuals look different but are not essentially different. Some genes have evolved extensively, whereas others have not. There are differences among loci; when each locus has been in the same population, the demographic features are the same of the population, the level of gene flow, the population size, and so forth, but they also are independent (*not clear to me*). Dr. Shriver noted that F_{ST} is a way to measure genetic distance between two sample sets. There can be a maximum F_{ST} of one if the population is totally different in allele frequency and a minimum of zero if the frequency is the same, but often it is somewhere in between. The X chromosome has a higher average F_{ST} than the autosome, illustrating that the X chromosome has experienced more evolution. There is a smaller effect for population size and every male is deficient for one X chromosome; therefore, a disease that is recessive in females is dominant in males, which initiates natural selection and more evolution. Dr. Shriver provided several examples, including that of the Duffy locus, which is fixed in West Africa for high frequencies and provides immunity to *Plasmodium vivax* malaria but is not found outside of Africa except by admixture. For this reason, it serves as a good ancestry informative marker (AIM) for measuring the admixture level among African American populations.

A test called Euro 1.0, based on 320 AIMs selected for European ancestor information, was developed to determine if there is stratification across Europe. Both the STRUCTURE and the principal coordinate plots of the marker panels show stratification patterns across Europe. Using a standard European American sample, he indicated that the principal coordinate plot speaks a little better to STRUCTURE in some ways, at least by revealing more about the genetic variation among the individual variation than STRUCTURE does. The model showed microclustering, itemizing Spanish (including Valencian and Basque), German, Jewish, French, and Italian ancestries. In terms of European AIMs, Dr. Shriver's group further screened them and measured some of the phenotypes that demonstrate that one can adjust for European stratification, including facial features and eye color genes. There is clearly facial and skull variation across

Europe, as other researchers also have noted. Dr. Shriver's group has also typed several African populations, including the Burungi (East Africa), Pygmy (central forests), Coisson (West Africa), and Bantu (southern Africa).

He continued with a query about African American origins within Africa, pointing out in a model that African Americans and West African parental populations cluster nicely. Because there is an inherent European American admixture in the African American genome complement, European African information content was reduced by removing markers that are informative across that particular axis. The results yielded a reasonable clustering of the West African groups together and suggested a geographic intersection of the breadth of the West African population. Further studies of African American origins should focus on all of Africa and not limited to West Africa. Dr. Shriver's group has 500 SNPs analyses on four indigenous American populations: Imer (Peru), Katchla (Bolivia), Mayans (Guatemala), and Nala (Mexico). These data exist and AIMs can be drawn from these to look at within and among variation. Indigenous Americans have been left out of most of the sequencing and allele frequency efforts, largely for political reasons. A future study could compare other Native American groups against these four populations. Finally, Dr. Shriver shared details of a study of how people see faces, which involved pictures a group of 75 individuals that had been collected in a study of human pigmentation. The study found that many facial features, not just skin color, can identify racial origins. This illustrates the coevolution between how one appears and how one can see people.

Genetic Epidemiology: The Value of Population Differences

Maria Elena Martinez, Ph.D., University of Arizona, Tucson, AZ

Population differences, such as race and ethnicity, disease and phenotype, allelic variation, lifestyle and environmental, cultural, and socioeconomic status, or combinations of these, are of great value to genetic epidemiologists and also may be useful in determining reasons for differences in cancer rates between different populations.

Characteristics of the U.S. Hispanic population changed dramatically between pre-1970 and 1990 to 2000. Pre-1970, the percent distribution of foreign-born Hispanics who entered the United States was 10.2, but between 1990 and 2000, that number increased to 45.8. Although most Hispanic populations increased their migration since 1970, there was a remarked drop in immigrants from Cuba from 38.6 percent in pre-1970 years to 28.4 percent by the 1990s. By far, the largest Hispanic population in the United States came from Mexico (59.3%), with most of the Mexicans residing in California (46.3%) and Texas (21.3%).

In the Americas, the countries with the highest breast cancer death rates (in 2000) per 100,000 included: Argentina (20.65), Canada (18.24), the United States (17.56), Cuba (14.82), Puerto Rico (13.75), Venezuela (13.34), Chile (12.56), Brazil (12.45), and Costa Rica (11.85). Rates for Mexico, Colombia, and Ecuador were the next highest but did not reach double digits. Using age-adjusted rates per 100,000, for Hispanics the female breast cancer incidence and death rates in the United States (1998-2000) were 89.8 (incidence) and 46.7 (mortality) compared to 141.1 (incidence) and 25.9 (mortality) for non-Hispanic Whites. Non-Hispanic Whites had a lower proportion of women diagnosed with breast cancer under the age of 50 compared to Hispanics in the US. Data comparing female breast cancer in Arizona by age at diagnosis and race/ethnicity to

Jalisco and Sonora, Mexico, showed that more women in Mexico were diagnosed at an earlier age (under 50) than non-Hispanic Whites in the US. Breast cancer death rates in Mexico from 1970 to 2000 have increased, particularly among younger women (30 to 64 years of age).

A binational comparative study of breast cancers and their risk factors among Mexican women in Mexico and in the United States is in the planning stages. In Mexico, breast cancer is the second cause of cancer death, after cervical cancer; however, recently it became the number one cause in more industrialized regions of the country. Data indicate that mortality rates in Sonora and other northern regions as well as more industrialized states of Mexico (e.g., Guadalajara) are higher than those in rural and southern states. Data furthermore suggest an early age of onset and later stage disease among Mexican women.

This study aims to: (1) compare profiles of tumor markers of prognostic and/or predictive clinical importance (ER, PR, HER-2/neu, Ki67) between women in Mexico and Mexican American women; (2) compare profiles of more novel tumor markers (p27, p53, cyclin E, PTEN, basal cytokeratins [5, 6, 17, 14], TGF beta 1) between women in Mexico and Mexican American women; and (3) assess whether differences in markers are more pronounced in postmenopausal women compared to premenopausal women and whether these are explained by factors associated with acquisition of lifestyles more representative of the United States (low parity, late age at first birth, adult weight gain pattern, and body composition such as waist circumference and body mass index). The study hypothesizes that the contribution of ancestral genes may differentially influence susceptibility to breast cancer risk and/or have more pronounced effects on specific disease subtypes. To this end, the study will: (1) assess the role of population mixing as a determinant of breast cancer susceptibility in the Mexican population, and (2) assess whether genetic markers of admixture segregate with the risk for specific subtypes of breast cancers among Mexican women. The study is limited by the need for genetic platforms for panels of genetic markers applicable to Hispanic populations. Additionally, almost 80 percent of the Mexican population considers itself mestizo, with different proportions of indigenous and European ancestry and an African component.

The importance of the proposed binational studies of breast cancer was summarized. Hispanics (especially Mexican Americans) are the largest growing minority population in the United States and represent a population that is largely underserved and underrepresented in research studies and clinical trials. Moreover, the population in the United States represents an unstable, highly migratory population with heterogeneous exposures compared to a stable population in Mexico with similar genetic background. Conducting studies of Mexican women in the United States and those residing in Mexico has the potential to help understand the etiology of disease for this population. This collaboration can help to address the question of whether migrants to United States are “different” than those in country of origin and in what ways they differ.

A Network of Investigator Networks in Human Genome Epidemiology

Teri Manolio, M.D., Ph.D., National Heart, Lung, and Blood Institute, Bethesda, MD

Data relating sequence variation to disease are accumulating exponentially, but identifying genetic determinants of complex diseases is hindered by a proliferation of small, poorly-designed and underpowered “convenience” studies that may also have biases in analysis and interpretation; selective reporting of positive results; lack of

standardization among studies; poor reporting of results; and difficulties in assessing environmental modification. Discordance in studies of associations of genetic variants may be caused by sampling errors or random type I errors in positive studies, lack of power in negative studies, genetic heterogeneity, population stratification or confounding, and differences in measurement methods. All of these problems can occur in many types of studies and highlight the need for coordination and collaboration.

Many genetic association studies are conducted using cases and controls of unclear origin with less than optimal data collection methods. Case control studies are difficult to conduct and may be best when nested within cohort studies to allow prospective collection of exposure information. Phenotyping of large cohorts is more difficult and expensive than genotyping and already has been done extensively in many existing studies. These studies should be brought together for association studies, but existing cohorts may not provide sufficient breadth, sophistication, or standardization of phenotyping or exposure information. Genotype prevalence, gene-disease association, gene-gene interactions, gene-environment interactions, and assessing genetic tests are key factors to be included in these types of studies.

Efforts are needed to bring together geneticists and epidemiologists. Creation of the Human Genome Epidemiology Network (HuGENet) has been envisioned as a global collaboration of individuals and organizations to assess the population impact of genomics on population health. The main components of this effort are information exchange and dissemination, training and technical assistance, and knowledge base development. HuGENet (www.cdc.gov/genomics/hugenet/default.htm) currently encompasses four coordinating centers around the world, eight collaborating journals, and more than 700 members from more than 40 countries; membership is free.

Another networking effort, the Public Population Project in Genomics (P3G Consortium; www.p3gconsortium.org), is a not-for-profit international consortium that promotes collaboration among researchers in the field of population genomics. Its mission is to provide the international population genomics community with the resources, tools, and knowledge to facilitate data management for improved methods of knowledge transfer and sharing and to create an open, public, and accessible knowledge database.

Proposed solutions to problems such as unavailability of data from population studies and publication bias include upfront study registration, which has been adopted for randomized clinical trials in databases such as ClinicalTrials.gov (www.clinicaltrials.gov), as a means to minimize publication and reporting biases and maximize transparency. For molecular research, however, upfront public registration of all ideas contradicts the individualistic spirit of discovery; instead, registration of investigators and data specimen collections is suggested.

Registries of data/sample collections might include networks of investigators working on the same disease, sets of genes, or field and could promote better methods and standardization while providing research freedom for individual participating teams. Such registries would permit thorough and unbiased testing of proposed hypotheses with promising preliminary data on large-scale, comprehensive databases and give due credit to investigators both for “positive” and “negative” findings. Registries of teams also could be created. A core registry should comprise information on the teams that already participate in a network. A wider registry also should record other teams working in the same field. Depending on the structure and funding opportunities of the

existing networks, additional teams may be allowed to join formally or at least be recorded to provide a more complete picture of the field. In addition, networks may have qualitative or other prerequisites for team membership. Central guidance and sharing of experiences also may be useful.

A Network of Networks could communicate and share expertise in statistical analytical methods, laboratory techniques, practical procedures, and logistics; coordinate and facilitate registries to avoid overlap; maximize efficiency and standardize methods and procedures; maintain an electronic list of registries that contain information on participating and nonparticipating teams; and compile an encyclopedia of validated molecular information for the disease or field. More than 20 international networks and registries currently exist and involve thousands of participants.

Steps and action items for a proposed Roadmap to facilitate Human Genome Epidemiology include developing a network of investigator networks to facilitate the remaining steps; improving study conduct, reporting, and harmonization; capturing published and unpublished data; improving data synthesis methods; and capturing and appraising evidence on the evolving “big picture” of a field. All of these steps are feasible and could be accomplished by the groups represented at this meeting.

A framework for risk evaluation in genetic association studies could be created by beginning with single teams and single studies reporting their results as either published or unpublished data; these can then be synthesized into systematic reviews and meta-analyses, which can be graded and synthesized and may result in field-wide synopses; these synopses can result in feedback to the individual teams that then conduct further research. HuGENet and the Network of Networks could facilitate this process by bringing the teams and studies together into systematic reviews; P3G and similar groups could help standardize protocols and methods and bring together published and unpublished data.

Priorities for connecting networks for common purposes should emphasize sharing of protocols and data and ensuring that a core of phenotypic and exposure information is collected in exchangeable formats using standardized methods. Another priority would be to genotype and correlate a core set of known variants and genome-wide markers across studies.

Charting the Iberian Peninsula Contribution to Ancestry in Latin America

Angel Carracedo, Ph.D., Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
Antonio Salas, Ph.D., University of Santiago de Compostela, Santiago de Compostela, Spain

The history of the Iberian Peninsula starts with a Neolithic group that settled in what today is Portugal and Spain in the second millennium BC. Following a Neolithic diffusion in the area, with different characteristics showing-up in different areas (e.g., Galicia in Spain and the Castrol culture in Portugal), the Roman Empire arrived in Spain during the second century AD, bringing their culture and dividing the region into several administrative provinces. Although the Roman Empire was important from a cultural and economic standpoint, it was not significant demographically. When the Romans arrived to the Iberian Peninsula, they found an area with many tribes embracing different cultures and languages. Based on linguistics, Spain could be clearly divided in two distinct areas (Celtic-speaking vs Iberian-speaking), although with some overlap where a

mixture of these two languages was spoken; there were no Basque-speakers at that time. Various Goth tribes arrived between the fourth and sixth centuries, dividing Spain into three parts, and the Arabs arrived in southern Spain between the eighth and 15th centuries. During the Medieval Ages, another Latin-derived language called Catalanian developed and expanded rapidly through most of the Iberian Peninsula. By the 16th century, Spain's languages began shaping into present linguistic dialects, with differences between the Basques, Galicians, and Catalans.

The Spanish immigration to the Americas, mostly from Central and southern Spain, was particularly important during the 16th and 17th century, with an upswing in the 18th century. The proportion of females to males was low in the 16th century but was almost equal by the 18th century. In addition, there was a dramatic decrease of Native Americans, especially in Peru and Brazil, during the 15th through the 17th centuries, likely caused by microbial infections. By the 18th century, the population in the Americas began to rise, and the annual growth rate was much higher than in most European countries. Immigration in the 19th and 20th centuries was mainly from northern Spain.

Dr. Carracedo observed that the Iberian population is not genetically heterogeneous. He described an ancestry analysis in which multiple questions were asked: (1) What is the level of population stratification in the Iberian Peninsula? (2) What is the level of population stratification that could have implications in population-based studies (type 1 error)? (3) How could this level of stratification affect the distribution of genetic variability of Iberian descents in Latin America? (4) Are there many disease or neutral markers observed in America that can be traced back to Iberia? To explore ancestries, two types of markers were examined: disease and (seeming) neutral markers (e.g., mtDNA [mtDNA], Y-chromosome SNPs, and autosomal SNP). Since Galicia is a relatively isolated region of Spain, the study proposed that genetic differences found in Galicia compared to other regions of Spain could be caused by a founder effects. Thus, for instance, there is high frequency of breast cancer (*BRCA*) gene mutations and doubled incidence of colorectal cancer in Galicia compared to the rest of Spain. These incidences may not have genetic causes, but the thesis matches well with the migration route.

An analysis of *BRCA1* and *BRCA2* genes in breast and ovarian cancer patients showed a substantial relation to mutations unique to Spain and evidence of founder effects. Additional disease markers studied included the adenomatous polyposis coli gene (colorectal cancer); the apolipoprotein B *R3500Q* gene (familial hypercholesterolemia); the *ABCC8* gene (hyperinsulinism of infancy); and the *HFE* gene (hemochromatosis). On the other hand, the analysis of Y-chromosome polymorphisms in samples from northern Africa and Spain were compared indicating the existence of micro-geographical differentiation in northern Iberia. Additionally, the mtDNA haplotype H, to give an example, was genetically dissected to confirm the existence of Iberian founder effects.

In conclusion, Dr. Carracedo noted that the Iberian Peninsula contains populations with varied cultural and genetic backgrounds, and its demographic contribution to the American genetic pool was significant, involving two main patterns of migration, the first from Central and southern Spain and the second from the north and northwest. Although intense gene flow occurs among regions, there is strong evidence that population stratification could have implications in association studies, including those using U.S. 'Hispanic' samples, as different Iberian origins may lead to an increase of false positives attributable to stratification.

Finally, a Spanish National Genotyping Center has been founded in Santiago de Compostela (Galicia, Spain) and encompasses different platforms for genotyping

(including pre- and post-genotyping) with the main aim of given support to high-throughput genotyping projects on e.g. complex diseases studies. It also houses a national DNA bank, which has made samples available for more than 55 projects, most of them related to cancer.

Patterns of Genetic Variation in Indigenous Populations From Mexico, Central America, and the Caribbean

Carolina Bonilla, Ph.D., Ohio State University, Columbus, OH

Dr. Bonilla addressed patterns of genetic variation in populations from Mexico, Central America, and the Caribbean. The objectives included: (1) a brief overview of the characteristics of each region and of published genetic studies conducted in the area, (2) commentary on the results obtained from research in some of these populations, and (3) an evaluation of what needs to be done to obtain a better genetic picture of the region.

The European conquest and colonization of America brought together continental populations that had been isolated for a long time. These were the original inhabitants of the continent, i.e. the indigenous Americans, the European colonizers (who were initially Spanish but were soon followed by other Europeans), and West Africans who were forcibly brought to the New World to provide labor. The way these populations interacted had important consequences for the populations of today.

Mexico

Mexico's population (~107 million people) is approximately 12 percent indigenous with about 7 percent indigenous language speakers of which 17% are monolingual. There is great linguistic diversity in Mexico, with 5 linguistic families but over 60 linguistic groups. The ethnic composition of Mexico consists of mestizos (60%), Amerindians (30%), Whites (9%), and others (1%) (CIA factbook). Mexico's ancestral populations include Mesoamerican cultures, European settlers and enslaved Africans. In Mesoamerica, major and minor state-like civilizations, some with large urban settlements could be found. The migration of Europeans to Mexico started after 1521 comprising primarily Spanish (mainly from the regions of Castilla, Andalucia and Extremadura), Jews, French and Italians. Enslaved Africans originated from West Africa for the most part, especially from Guinea, Senegambia, Angola, and Congo.

These populations gave rise to a mixed group of individuals, called mestizos. Mestizos originated as a result of Spanish and Native American admixture. The definition of mestizo was provided by the National Institute of Anthropology as a person who is born in Mexico, has a Spanish-derived last name, and has at least three generations of Mexican ancestors (Gorodezky et al., 2001).

Several genetic diversity and admixture studies have been conducted in Mexico that examined indigenous and mestizo groups for autosomal DNA and protein polymorphisms such as blood groups and serum proteins, histocompatibility antigens, variable number of tandem repeats (VNTRs), and short tandem repeats (STRs) (see papers by Buentello-Mallo et al., 2003; Cerda-Flores and colleagues; and Lisker and colleagues). On the other hand, fewer studies have been conducted on mitochondrial DNA (mtDNA) and Y-chromosome markers (Torrioni et al., 1994; Green et al., 2000).

The analysis of ancestral proportions in Mexican indigenous and mestizo populations has shown a significant Native American contribution, widespread European ancestry, and a considerably smaller or even absent West African ancestry. However, there is great variation in admixture proportions among Mexico's regional populations. European ancestry is greater in the North than in the rest of the country, whereas West African ancestry increases towards the coastal areas with a concomitant descent of Native American ancestry.

We have studied a rural population from the city of Tlapa in the state of Guerrero, which lies on the Pacific coast (Bonilla et al., 2005). Tlapa, however, is located amid mountains on the eastern part of the state. Individuals in Tlapa belonged to three ethnicities: Nahua, Mixtec and Tlapanec. Individuals of mixed ethnicities and self-reported mestizos were also included in the sample. A total of 24 autosomal ancestry informative markers (AIMs); the four typical Native American mtDNA haplogroups; and Y-chromosome marker DYS199 C/T, were examined.

The Native American DYS199*T allele frequency was high in all native groups with some variation. The mtDNA haplogroups were overwhelmingly Native American in origin even among mestizos. Among mtDNA lineages, haplogroups A and B were the most frequent in all groups, while haplogroup D exhibited the lowest frequency. The admixture estimates based on the 24 autosomal AIMS showed Native American ancestry to be very high in the population of Tlapa (~94%).

An examination of population stratification showed that there was evidence of genetic structure in the population of Tlapa when mestizos were part of the sample but that was not the case when mestizos were not included.

Central America

Central America consists of seven countries, Belize, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica and Panama, with a total population of about 40 million. Indigenous populations represent up to 44 percent of the population of each country, with the highest indigenous population residing in Guatemala.

There have been few studies conducted about Central American genetic diversity or admixture, with the exception of Costa Rica. Most of the studies have been performed for forensic purposes using STRs and VNTRs (e.g., the Combined DNA Index System [CODIS]).

Costa Rica was the point of contact between Mesoamerican and South American cultures. At present there are eight indigenous groups, which represent approximately 1 percent of the population. The country was colonized by Spain in 1561 and later received other European migrants like Italians, Germans and Jews. The African influence is highest on the Atlantic coast, where the slave trade was concentrated.

Genetic studies in Costa Rica have estimated the degree of admixture in mestizo individuals and have also looked at affinities between indigenous groups, using mostly classical markers (Barrantes, 1993; Azofeifa et al., 2001; Ruiz-Narvaez et al., 2005). The Cabecar are a less acculturated and admixed group, and the Huetar show higher European admixture and higher Y-chromosome diversity. Analyses of mtDNA and Y-

chromosome diversity in the Chibchan tribes have found a similar population structure for both systems, which indicates that it is likely that there was no difference in the migration rates of males and females. In addition, the origin of the Chibchan group has been dated as occurring 7,000 to 10,000 years BP, using coalescent estimates based on uniparental markers.

Among mestizos, admixture analyses have estimated parental contributions as 61% European, 30% Native American and 9% West African, on average (Morera et al., 2003). There is, however, regional variation. For example, there is greater European ancestry in northern and central Costa Rica, greater Native American ancestry in southern Costa Rica, and greater African ancestry along the coasts (Madrigal et al., 2001). Studies of mtDNA found that 83 percent of the population had a Native American maternal lineage in the Central Valley, whereas only 5% of paternal lineages were indigenous (Carvajal-Carmona et al., 2003).

Regional variation also seems to be the case in other Central American countries such as Nicaragua or Guatemala, however, more studies need to be conducted in these nations to obtain a more complete picture of their genetic make-up.

Studies of mtDNA haplogroup frequencies in Mexico and Central America have shown a high frequency of haplogroup A across the region and somewhat less of haplogroup B. There were much lower frequencies of haplogroup C; haplogroup D was almost absent. Other haplogroups, probably introduced by admixture with Europeans and/or Africans, are almost non-existent among indigenous groups but the Maya, and are present in small proportions in the mestizo populations of Mexico and Costa Rica. DYS199*T frequencies also were examined and were highest in the Mixtecs-Guerrero group (and over 50% in all indigenous populations) and lowest in the mestizos of Central Valley of Costa Rica, probably because of admixture of native women with non-indigenous men.

The Caribbean

The Caribbean islands can be classified according to size in Greater and Lesser Antilles and the Bahamas. They can also be classified based on which European nation colonized the area, in Spanish, British, French, Dutch and Danish West Indies.

Studies in the Caribbean have concentrated primarily on the Spanish Caribbean. Estimates of ancestral proportions were obtained for Puerto Rico using mtDNA data (Martinez-Cruzado et al., 2001; 2005), and autosomal classical markers (Hanis et al., 1991), AIMs (Bonilla et al., 2004; Salari et al., 2005), and STRs (Zuñiga et al., 2006). Population samples studied by these researchers included Puerto Ricans living in Puerto Rico and Puerto Ricans who had migrated to the US.

Data on Cuba and the Dominican Republic is not as abundant as on Puerto Rico. Analyses of mtDNA and classical autosomal markers have been published for Cuba (Hanis et al., 1991; Torroni et al., 1995). Within the Dominican Republic, a study on mtDNA and diabetes reported that the controls had ~ 52% of Native American ancestry which was higher than that of cases (Tajima et al., 2004).

Populations that are part of the non-Spanish Caribbean exhibit very high West African ancestry with almost negligible Native American contribution with the exception of

Trinidad (Molokhia et al., 2003; Miljkovic-Gacic et al., 2005), while populations in the Spanish West Indies are more trihybrid with significant contributions of all ancestors but with a major fraction of European ancestry. The differences observed are most likely due to the way different European nations conquered and colonized the areas in question. British, French, Dutch, and Danish colonies saw the rapid decline of their native populations and received a massive slave trade because of their plantation economy, something that did not occur in the Spanish colonies.

We have analyzed samples from Puerto Rico, Barbados, Jamaica, and St. Thomas, using a set of ~40 autosomal AIMs and estimated contributions from the three parental populations. European and Native American ancestry were highest in Puerto Rico whereas Barbados showed the highest levels of African ancestry and lowest levels of Native American ancestry.

We also tested for the presence of population structure due to admixture (admixture stratification) in these islands. Among the non-Spanish West Indies, in Barbados there is no evidence of stratification, Jamaica exhibits a low but nevertheless significant level of structure, and St. Thomas is the population that shows the largest degree of structure. So even though their levels of admixture are not significantly different, these populations do differ in the amount of genetic structure present in them. In Puerto Rico, on the other hand, there is extensive admixture stratification.

Several conclusions can be drawn from published data and our research, including that there is significant heterogeneity in Mexico, Central America, and the Caribbean that could be explained in part by differences in admixture patterns. In addition, the coastal areas of Mexico and Central America exhibit higher West African ancestry than inner continental areas. West African ancestry is predominant in the non-Spanish Caribbean, whereas Native American ancestry is high in Mexico. Moreover, all indigenous groups show some level of nonindigenous ancestry. Similar to what is seen across Latin America, there is evidence of sex-biased gene flow in these populations. An important point that stems from these findings is that populations with similar ancestral proportions may differ in population structure. The existence of genetic structure within a population may have important implications for the successful mapping of complex disease/trait genes in that population.

Considerations for future work include: focusing on understudied populations such as the Dominican Republic, El Salvador, Guatemala, and Haiti; estimating ancestral proportions using larger and more informative sets of markers; examining admixture stratification; extending work on uniparental markers; and creating a database that compiles genetic information on all Latin American populations.

Genetic Consequences of the Recent African Diaspora

Rick Kittles, Ph.D., The University of Chicago, Chicago, IL

Africans arriving to the Americas during the time of the Transatlantic Slave Trade originated primarily from West and West Central Africa and to a much lesser extent from East Africa. DNA analysis, particularly mtDNA and Y-chromosomal DNA, can be used to track the Diaspora and provide insightful information about migration patterns.

The genetics of African-descent populations in the Americas have to be placed in a historical, sociopolitical, and psychological context in order to understand self-identified ethnicity (SIRE). Clearly SIRE varies across the Americas among people with African ancestry depending on the social/political histories of individual communities. Defining individuals as Black American, Caribbean, or African American differs according to the locale. In the United States, African Americans have been legally and socially defined by the “one-drop” rule, a legislated rule that, during the period of slavery, classified a group of people based on having at least one ancestor of African descent. This social definition classified people regardless of mixed ancestry as “Black”. Thus, African Americans today represent a large, heterogeneous “macro-ethnic” group with diverse genetic ancestries. Interestingly, Hispanic/Latino populations are also highly heterogeneous due to a mixture of high proportions of Native American, European, and in some communities African ancestry.

The Transatlantic slave trade, occurred from the early 1600s to the 1800s. Currently, attempts are underway to examine the genetic, health, social, and political implications of this forced migration. During the Middle Passage, tens of millions of enslaved Africans were brought to the Americas, but not all survived, which may have implications for health issues for African-descended communities. As an example, prostate cancer incidence and mortality data based on data from NCI’s Surveillance, Epidemiology, and End Results (SEER) Cancer Registries and the International Agency for Research on Cancer shows that in North America and the Caribbean (Puerto Rico, Dominican Republic, and Trinidad and Tobago) populations with high African ancestry also have high incidences of mortality from prostate cancer.

Approximately 95 percent of enslaved Africans came from West and Central Africa, as determined by shipping and naval records. This information is useful for understanding the genetic features of these African-descent communities. In the African American population, most of the genes of European ancestry are derived from European men. This is largely due to the behavior of slaveholders in the antebellum south (7-10 generations ago) and is evident in variation for sex-linked markers among African Americans. This recent admixture appears to differ geographically across the Americas and resulted in increased linkage disequilibrium (LD) in African Americans which can be useful for gene mapping. A significant amount of variation can be seen in mtDNA and Y chromosomes in West Africa, as well as a variety of different haplotypes that also are found in African Americans. Analysis of West African mtDNA variation across 15 populations found significant correlations between genetic variation and geographic distance but not language.

Studies of different African proportions in African American communities have used sex-linked markers, such as Y chromosomes and mtDNA, to ascertain regional variation in African Americans. The Black Rice hypothesis posits a regional preference of enslaved Africans, based on the principal cash crop in those plantations, which may have led to regional stratification of the African American gene pool. In South Carolina, the principal crop for most of the antebellum period was rice, which led plantation owners in South Carolina to prefer enslaved Africans from regions of West Africa, where the inhabitants had considerable expertise in rice cultivation. The historical, linguistic, and cultural studies suggest continuity between these regions (Sierra Leone, Liberia, and Guinea) and African Americans from South Carolina. Another analysis of Y-chromosome variation in men of African descent in the District of Columbia, South Carolina, Jamaica, and St. Thomas islands showed significant variation between populations for Y

chromosome genetic markers. Approximately 30 to 40 percent of the Y chromosomes in men of African descent were of European ancestry. Analysis of mtDNA from 4,000 African Americans revealed that a significant proportion of African American maternal lineages (~36%) originate from regions of Africa historically known for grain cultivation (i.e., Senegambia, Sierra Leone, and Liberia). Y chromosome analyses suggest that about 15% of paternal lineages in some of the U.S. south (Georgia, Virginia, and Louisiana) trace to Angola. Approximately 50 percent of the Y chromosomes common in the area of present-day Liberia are found in the Mississippi area of the U.S. south. Very few (<5%) Native American maternal and paternal lineages have been found in African American communities or in communities throughout Central and South Americas.

Many autosomal markers are useful for estimating ancestry and can reveal information about structure in communities. An analysis using the structure program for 112 ancestry informative markers focused on West Africans from Cameroon, European Americans, and African Americans from Washington, DC, found a significant amount of population substructure in the African American population. Several trends have been discerned concerning the distribution of European admixture or genetic ancestry in African American communities across the United States. African Americans living in the urban North have a higher percentage of European admixture than do those in the rural South (with the exception of New Orleans and other cities in Louisiana), as do African Americans living in the western United States, particularly Washington State and northern California. Another study showed that St. Louis Valley Hispanics had significant amounts of Native American and European ancestry but little West African ancestry while Puerto Ricans have higher African ancestry. African Caribbean populations vary also, for example, Jamaicans have higher Native American ancestry than do people from Barbados and St. Thomas.

Although the *New England Journal of Medicine* (January 2006) considered self-reported race to be accurate enough to be included in genetic studies, race can be confounded by biology, environment, diet, and lifestyle. Genetic ancestry as a proxy for genetic background or for disease susceptibility also must be used with care. This is because confounding variables such as racism and SES appear to be correlated with genetic ancestry in some communities.

Admixture in Hispanic/Latino Populations: Distribution of Ancestral Population Contributions in the Continental United States

Ranajit Chakraborty, Ph.D., University of Cincinnati, Cincinnati, OH

Dr. Chakraborty presented information on the distribution of the contribution of parental population in the United States, particularly the Hispanic population, which generally refers to the people or culture of Spain and Portugal. The ethnic category evolved from a decision by the U.S. Office of Management and Budget (OMB) in 1978, which stated that "a person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race" was to be described as a Hispanic (*Federal Register*, Washington, DC, 1978, vol. 43). The 2000 U.S. Census categorizes Hispanic and Latino groups as Mexican, Puerto Rican, Cuban, and "other Hispanic/Latino," which includes Dominican, Central American, South American and others; within those subgroups are the countries of origin. The United States has approximately 40 million persons who fit the description of Hispanic, two-thirds of which are of Mexican origin.

The composition of the Hispanic population differs considerably within the four regions used by the U.S. Census Bureau (i.e., Northeast, Midwest, South, and West). The implication of these differences is that if a group is defined as Hispanic from a specific region, then the genetic composition of the Hispanic population will differ from Hispanic populations found in other regions. For example, in the Midwest, Hispanic population is approximately 70 percent Mexican, but in the Northeast, Mexicans comprise only approximately 10 percent of the Hispanic population. Age composition is an additional factor to consider in population differences in the context of complex diseases. If pediatric or early onset diseases are examined, approximately one-fourth of Central American or Cuban populations are 18 years or older; Puerto Ricans and Mexicans, in contrast, are younger.

There is a "Hispanic Paradox" found in studies of Hispanics in the United States, that indicate that they have better or similar health compared to that of non-Hispanic Whites despite lower incomes and less education. This population also has lower mortality compared to non-Hispanic Whites, although these observations are contested by some researchers. There are two hypotheses to explain the Hispanic Paradox. One is the "Healthy Migration Effect," which states that only healthy persons move and migrate from their country of origin. The other hypothesis is the "Salmon Hypothesis," which states that sick people tend to migrate back to their country of origin. Although there are many advocates for these theories, the observed paradox is not totally explained by either of these hypotheses. There are possible biological or genetic effects as well.

When the history of admixture studies are examined, they can be grouped into the following three categories: admixture at group level using genomic markers; admixture at the individual level, which began in the 1970s; and admixture components revealed at the mtDNA and Y-chromosome level to detect gender-biased contributions of ancestral populations. Each of these groups provides another dimension to the implications of admixture studies. Beyond the geographic-political boundaries of the United States, such as in Mexico, there are also such differences. Studies in Puerto Rico and Cuba show that contributions from ancestral populations differ among studies.

In a forensic study using DNA, performed by a graduate student of Dr. Chakraborty, groups from five regions of the United States were assessed for allele or genotype frequencies for specific DNA loci. Results from analysis of molecular variance for ethnic differences considered two groups: West (California, Nevada, and Southwest) and East (Florida, New Jersey, Pennsylvania, Virginia, and Southeast). Significant differences were found between the two groups and among populations within groups. If the admixture component is computed using standard measures taking current continental data from Europe, Africa, and Native Americans, there are differences among the mixture components between populations within groups, as well as between groups.

In a study using the 1990 Census data, researchers (Bertoni et al.) tabulated the proportion of persons of Mexican origin in the same populations. There were vast differences in the proportions, which corresponds to the ancestral origin of the groups in each region. Other types of data analysis involve Cubans. Dr. Chakraborty noted that the manner in which populations are placed in studies will affect the types of clustering that result. He showed that other individual admixture studies revealed the same results. In 1994, just before the International Congress of Genetics, mitochondrial diversity was being examined. Studies were initiated where the same populations also were subjected to admixture analysis by mtDNA. The admixture component coming from the

autosomal markers can be different from the mitochondrial. Similarly, for the same population, the admixture component for mitochondrial and Y-chromosome markers can be very different from the autosomal loci.

Dr. Chakraborty offered six conclusions:

- (1) Hispanic groups within the continental United States are heterogeneous by their country of origin and culture, as well as in their genetic composition.
- (2) Contributions of ancestral populations in Mexican Americans, Cubans, and Puerto Ricans are substantially different.
- (3) Because of unequal geographic distributions of different Hispanic groups in the United States, the genetic composition of Hispanics defined by geography alone may be even more problematic.
- (4) In all of these groups, gender-biased gene flow is evident.
- (5) AIMs provide better efficiency of admixture detection at both group and individual levels.
- (6) Phenotype-dependency of AIMs, however, may make them subjected to effects of natural selection, biasing admixture estimates derived from them.

From “Mestizo” to “Metis”: Insights and Perspectives on Admixture in Mexico and Canada

Esteban Parra, Ph.D., University of Toronto, Toronto, Ontario, Canada

Dr. Parra focused on an admixture study of a sample of type 2 diabetes (T2D) patients and controls from Mexico City and a brief review of admixture in Canada, with a special focus on the Métis population.

The admixture study in Mexico City enrolled 286 unrelated T2D patients and 276 unrelated controls from Mexico City. Samples corresponded to individuals affiliated with the Mexican Institute of Social Security, which serves approximately 50 percent of the Mexican population, and additional information (e.g., sex, age, body mass index, and education) was collected. Approximately 70 AIMs served as autosomal markers; mtDNA and Y chromosome polymorphisms also were studied. The study revealed that the average proportion of Native American ancestry was 65 percent, European was 30 percent, and West African was 5 percent. The average number of generations since admixture was seven generations. There also is strong evidence of sex-biased gene flow based on mtDNA and Y chromosome evidence. The test was based on posterior predictive check probability and showed strong evidence for the presence of genetic structure. This was reflected in a large number of associations between unlinked markers; 1,900 tests were conducted and 442 significant associations (23%) were discovered, although only 5 percent were expected. These results emphasize the need to control for population stratification when carrying out conventional association studies

Continuous gene flow and assortative mating help to maintain genetic structure, and when exploring the relationship between ancestry and education, the study found strong

evidence of socioeconomic stratification in this sample, which is an important social issue in Mexico as not everyone has the same access to education. Using a logistic regression model with education as an outcome, it was determined that people with 100 percent European ancestry are 2.4 times more likely to have higher education than people with 0 percent European ancestry. Furthermore, mating likely is not random with respect to socioeconomic status in Mexico, and socioeconomic status shows a strong association with ancestry. This is probably one of the major factors explaining the presence of genetic structure in this population.

The results of the present study indicate that the Mexican population is suitable for admixture mapping, both in terms of admixture proportions and the number of generations since admixture (related to mapping resolution). A genome-wide map of Native American/European AIMs will be available soon, opening the door to admixture mapping applications in many populations across the Americas.

In Canada, admixture is recognized by the government to include three aboriginal populations: North American Indians (also called First Nations), Métis, and Inuit. In the 2001 Canadian census, approximately 1 million people identified themselves as aboriginal: 62 percent self-reported as North American Indian, 30 percent as Métis, and 5 percent as Inuit.

There have been very few studies to characterize admixture in Canadian aboriginal groups. Szathmary, et al. (1983), used serum protein and red cell enzyme markers to estimate the European admixture in the Dogrib (Northwest Territories) at 8.7 percent. Field, et al. (1988), used immunoglobulin allotypes (GM and KM) to estimate between 12 and 20 percent European admixture in Haida and Bella Coola. Finally, the European haplogroups H and T have been observed among the Ojibwa from the Great Lakes region (Schurr, 2000).

The Métis population results from admixture between indigenous Canadians and Europeans and traces back to the initial colonization of Canada by the French and British. Their mixed ancestry is reflected in many aspects of their art, culture, and lifestyle. In the 2001 Canadian census, there was a 43 percent increase of self-reported Métis from the previous census and was the largest population gain of the Canadian aboriginal groups. More than two-thirds of the Métis live in urban areas, and no admixture studies have been carried out in the Métis population. Although few studies have characterized continental admixture in Canadian populations compared to the United States, Central and South America, and the Caribbean, admixture studies in the Métis could bring a better understanding of the history of this population. It also could have the potential to explain the reasons for the prevalent differences observed between European and Native American populations for some phenotypes and diseases.

Tag Single Nucleotide Polymorphisms (SNPs) in Admixed Populations

Eduardo Tarazona-Santos, Ph.D., Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Dr. Tarazona-Santos noted that many SNPs now are available, but admixed Latino American populations, including Mestizos and Native American populations, are underrepresented in studies that address human genome diversity. This underrepresentation results from cultural issues and logistical problems.

The lack of representation of Hispanics, Latin American, and Native American populations posits the problem of tagSNPs portability. For instance, if tagSNPs are obtained in the European population, it is important to determine how applicable these tagSNPs are to Latin American/Hispanic and Native and populations.

Latin American populations are typically tri-hybrid ones, and have received contributions from Native American, European, and African parental populations. Linkage disequilibrium in the admixed population, a determinant of tag-SNPs is a function of the average LD and the covariance of the allele frequencies in the parental populations. The process of admixture itself is quite complex but can be simplified for study purposes. Dr. Tarazona-Santos and colleagues used a simplified approach to test the patterns of LD across samples and the portability of tagSNPs for specific genes on Chromosome 22. In total, they analyzed 57 SNPs for six genes on Chromosome 22. They measured how frequently tag-SNPs ascertained in European populations, are portable to Native American and admixed populations with different degree of admixture. They concluded that tagSNPs ascertained in European populations were portable to Native American and to admixed bi-parental populations (Native American and European). Reduced tagSNPs portability was observed when African admixture is present.